“When Numbers Lie: Detecting Trend Breaks with Semantic Signals”

Agić T., Kurić A., Rodin V., Sivro K.

Univerzitet u Sarajevu
Elektrotehnički fakultet
Sarajevo

OBJECTIVES

Primary Goal: Improve short-term S&P 500 price prediction accuracy by combining financial sentiment analysis with technical indicators using advanced deep learning architectures.

Research Questions:

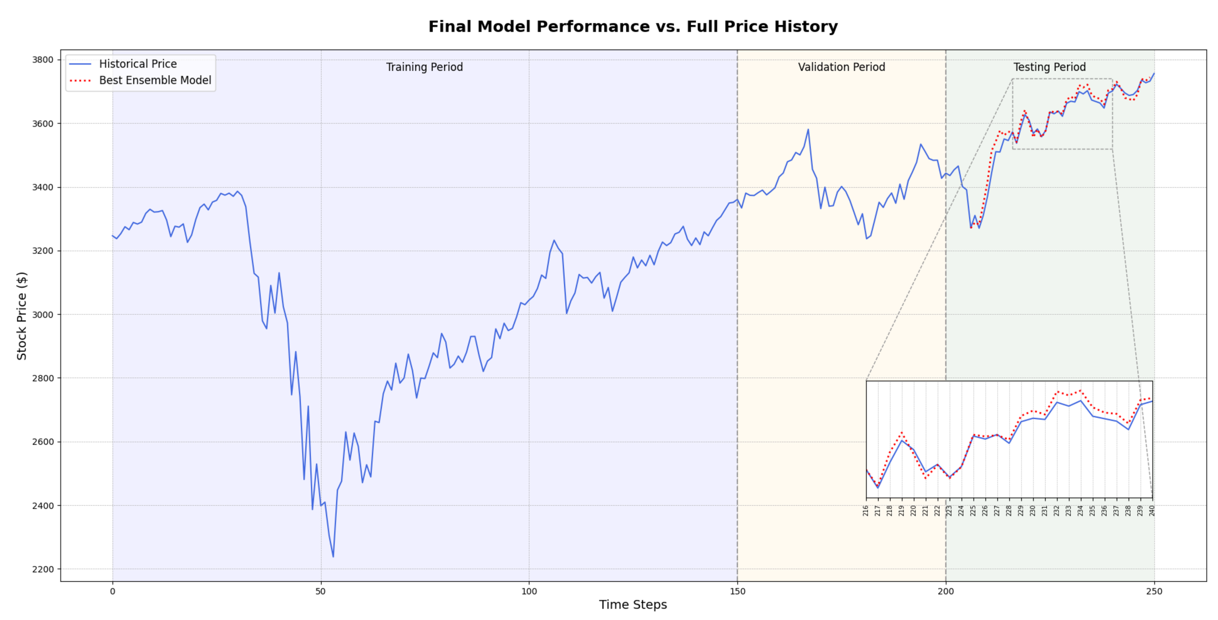
- Can social media sentiment enhance traditional time series forecasting?
- Which neural architecture best captures temporal patterns in financial data?
- How does ensemble methodology improve prediction stability?

Innovation: Novel integration of CNN-LSTM architecture with Twitter sentiment features and stability-weighted ensemble strategies.

KEY RESULTS

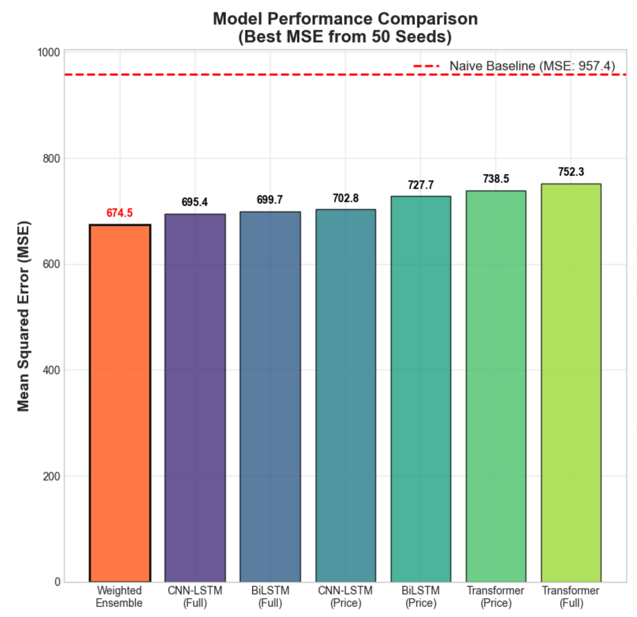
Best Performance Achieved:

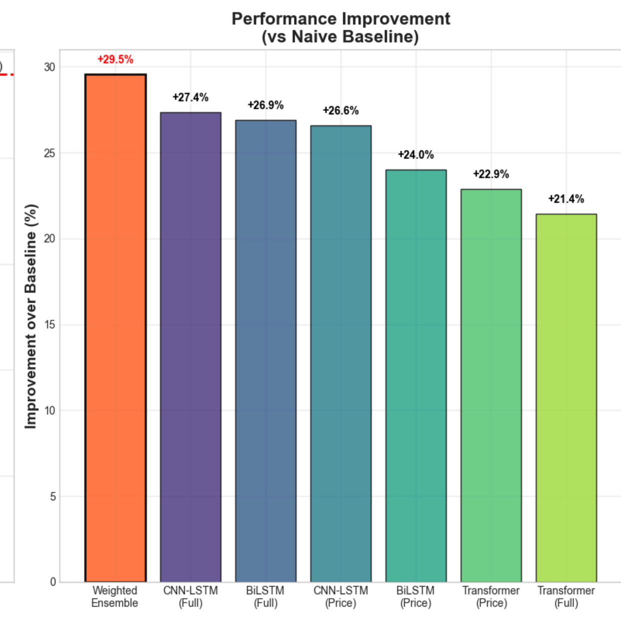
- Weighted Ensemble: 29.5% improvement (MSE: 674.5)
- CNN-LSTM (Full): 27.4% improvement (MSE: 695.4)
- BiLSTM (Full): 26.9% improvement (MSE: 699.7)



MULTI-SEED ANALYSIS RESULTS (50 random seeds)
Naive Baseline MSE: 957.41

Model	Best MSE	Mean±Std	Best Imp%	Stability	Range
CNN-LSTM (Full)	695.4	779.3±42.6	27.37%	0.945	157.7
BiLSTM (Full)	699.7	808.4±118.2	26.92%	0.854	686.5
CNN-LSTM (Price)	702.8	823.4±57.2	26.60%	0.930	346.9
BiLSTM (Price)	727.7	894.8±168.5	23.99%	0.812	1156.2
Transformer (Price)	738.5	1004.1±315.1	22.87%	0.686	1499.7
Transformer (Full)	752.3	981.5±206.2	21.42%	0.790	991.7





Key Findings: Full features outperform price-only configurations except for Transformers. CNN-LSTM achieves the best stability-performance trade-off, while weighted ensemble strategies consistently outperform individual models. Twitter sentiment provides measurable prediction improvements across most architectures.

CONCLUSIONS

Validated Hypotheses: Social media sentiment significantly enhances prediction accuracy, with CNN-LSTM effectively capturing both spatial and temporal patterns. Weighted ensemble strategies provide consistent performance gains, while multi-seed analysis proves crucial for reliable model selection.

Practical Implications: The 29.5% improvement translates to significant trading advantages, with stable models reducing deployment risk in volatile markets. The architecture supports potential real-time applications for low-latency prediction systems.

Impact: This research demonstrates that combining traditional technical analysis with modern sentiment data through advanced neural architectures can substantially improve financial forecasting accuracy while maintaining model stability.

FUTURE DIRECTIONS

Immediate Extensions:

- Multi-horizon predictions (5-day, 10-day forecasting)
- Alternative data integration (options flow, insider trading, macro indicators)
- Advanced architectures using graph neural networks for market relationships

Long-term Vision: Development of risk-adjusted optimization using Sharpe ratio and maximum drawdown objectives, extension to multi-asset frameworks, and production deployment with real-time trading systems featuring continuous learning capabilities.

Limitations: Current scope limited to single market regime (2020 data), computational requirements may challenge real-time deployment, and sentiment data may contain noise from non-financial discussions.

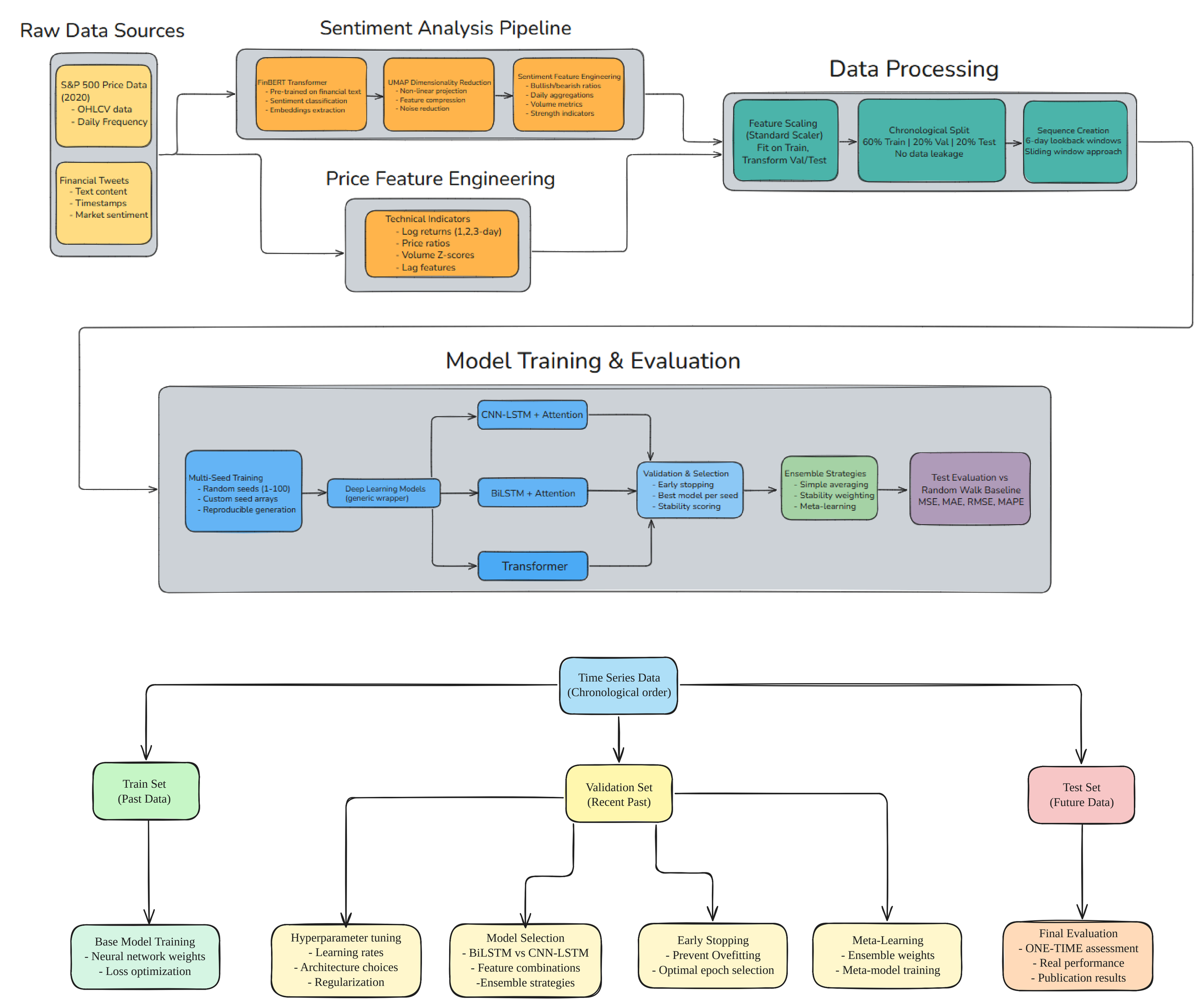
METHODOLOGY

Data Integration: The study combines S&P 500 daily OHLCV data from 2020 (250 trading days) with the StockEmotion Twitter dataset. We engineered 29 total features: 13 technical indicators including log returns, price ratios, and lag features, plus 16 sentiment metrics capturing bullish/bearish ratios and tweet volume indicators.

Model Architectures: Three advanced deep learning models were implemented:

- CNN-LSTM: Convolutional feature extraction + LSTM temporal modeling + Attention
- BiLSTM: Bidirectional LSTM with attention mechanisms
- Transformer: Multi-head self-attention with feed-forward layers

Evaluation Framework: Models were evaluated using chronological 60/20/20 splits with 50-seed stability analysis, measuring performance against a naive baseline using MSE and percentage improvement metrics.



TECHNICAL HIGHLIGHTS

Implementation Details: Built using PyTorch with CUDA acceleration, featuring Adam optimizer (lr=0.0015), early stopping, and dropout regularization. The system uses 6-day lookback windows with sliding approach and temporal attention mechanisms for improved interpretability.

Data Integrity Measures:

- Strict chronological splits with no data leakage
- StandardScaler fit only on training data
- Reproducible results with fixed base seed (42)

TECHNICAL SPECIFICATIONS:

Model Parameters: Hidden Units: 50, Layers: 2, Dropout: 0.15, Batch Size: 24, Max Epochs: 300, Patience: 14, Lookback Window: 6 days, Test Ratio: 20%

Performance Metrics: Baseline MSE: 957.4, Best Individual: 695.4 MSE (CNN-LSTM), Best Ensemble: 674.5 MSE (Weighted), Stability Range: 157.7 - 1499.7

REFERENCES

- **StockEmotions: Discover Investor Emotions for Financial Sentiment Analysis and Multivariate Time Series** [https://arxiv.org/abs/2301.09279]
- **SENN: Stock Ensemble-based Neural Network for Stock Prediction** [https://github.com/louisowen6/SENN]
- **Investigating Twitter Sentiment in Cryptocurrency Price Prediction** [https://github.com/BaharehAm/Cryptocurrency-Price-Prediction]